Zero Intelligence Models of the Continuous Double Auction: Empirical Evidence and Generalization

Martin Šmíd*

March 28, 2013

Abstract

The paper is devoted to statistical analysis of zero intelligence models of continuous double auction, in particular to an estimation of parameters of limit order books. We review "classical" zero intelligence models and show, by means of tick-by-tick quote (L1) data, their poor fit to liquid markets. Therefore, we define a generalized zero intelligence model which copes with the discrepancies found and devise a method of its estimation, which we show to be - up to a minor approximation - consistent and asymptotically normal. We demonstrate the model to at least mostly fit the data of three US stocks from US electronic markets.

Keywords: Continuous double auction, zero intelligence models, econometric estimation, consistency, asymptotic normality

1 Introduction

As most of the trading at today's securities markets is done according to the continuous double auction, the importance of mathematical modelling of this trading mechanism grows. Since it is probably impossible to build at least partially tractable model of a market with the continuous double auction

*Institute of Information Theory and Automation of the ASCR

assuming rationality of the agents, its existing models are “zero intelligence” (ZI), i.e., assuming a purely random arrival and (eventual) cancellation of the orders.

From a number of similar models of that kind, let us name Stigler (1964); Maslov (2000); Luckock (2003); Smith et al. (2003); Mike and Farmer (2008); Cont et al. (2010). All of those models assume unit order sizes, Poisson arrivals of market and limit orders with intensity dependent on the current best quotes and locally constant cancellation rates of waiting limit orders depending on the best quotes, too.[1] Some of the models assume continuous prices (Maslov (2000); Luckock (2003)) while the rest of them works with (more realistic) discrete (tick) prices.

As we already indicated, the distribution of even the simplest zero intelligence models cannot be expressed analytically; however, partial theoretical results are known (see Slanina (2001); Cont et al. (2010); Cont (2011); Šmíd (2012)) including the ergodicity of a discrete model with bounded prices (Cont et al. (2010)) and an analytic formula for the conditional distribution of the order books given a history of the best quotes applicable to the majority of known ZI models (Šmíd (2012)) which, when combined, are sufficient for a construction of consistent and asymptotically normal estimators of the parameters of the models - one of such estimators is presented by the present paper.

Even if the ZI models are claimed to be are able to mimic many stylized facts observed in reality, such as fat tails or the non-Gaussianity (Slanina (2001, 2008); Šmíd (2008); Cont (2011)) an ”old-style” econometrics has, as to the best knowledge of the author, never been applied to verify their accordance with reality. One of the reasons for this seems to be a difficult accessibility of the order book (L2) data. However, as we show in the present paper, L2 data are not necessary because a sample dependent on the parameters of the limit order book may be elicited even by means of quote (L1) data, even if only closest parts of order books may be estimated that way. In particular, we observe sizes of jumps of the quotes out of the spread and use the fact that the jumps greater than one imply emptiness of the first tick of the order book whose stochastic properties depend on the order placement and cancellation parameters.

We start our exposition by an introduction of a sufficiently general version

[1] The only exception is Mike and Farmer (2008) where the cancellation rate depends, in addition, on the order books’ size and imbalance.

of a ZI model, covering models of Cont et al. (2010), Stigler (1964), a discretized version of Luckock (2003) and slightly modified version of Smith et al. (2003) and we prove the ergodicity of the introduced model.

Consequently, we demonstrate the inconsistency of the ZI model with reality. In particular, we show how the empirical conditional probability of larger jumps of ask evidently contradicts its theoretical counterpart.[2]

To accommodate the discrepancies found, we propose a generalized zero intelligence (GZI) model. In particular, while we keep assuming Poisson order arrival and time-constant cancellation rates, we add the possibility of shifting of the orders, allow slowly cancellable (strategic) orders and admit multiple orders' placement at the times of the quotes' jumps. Moreover, we propose an estimator of the model which we show to be consistent and asymptotically normal up to a small approximation.

Consequently, we apply our estimator to L1 data of three US stocks traded at three US electronic markets, Our estimation is shown to bring at least partially significant results fir seven out of the nine stock-market pairs. Moreover as the parameters exclusive to the GZI model came out significant in six cases our empirical results falsifies the ZI model statistically correct way.

The paper is organized as follows: First, the ZI model is presented (Section 2) and the dataset we work with is introduced (Section 3). After a confrontation of the ZI model with the data (Section 4), our generalized model is formulated and its theoretical properties, necessary for the estimation, are stated (Section 5). Consequently, the estimation is performed and the results interpreted (Section 6). Finally, the paper is concluded (Section 7). Proof of the asymptotic properties of the NLS estimator is given in the Appendix.

2 A Zero Intelligence Model

Let us consider a general discrete zero intelligence model with unit order sizes described by a pure jump type process

$$\Xi_t = (A_t, B_t), \qquad t \geq 0,$$

where

$$A_t = (A_t^1, A_t^2, \dots, A_t^n)$$

[2]Even if this disagreement is shown only graphically, our later empirical results confirm it a statistically correct way.

and

$$B_t = (B_t^1, B_t^2, \dots, B_t^n)$$

are the the sell limit order book, buy limit order book, respectively. In particular, A_t^p (B_t^p) stands for the number of the sell (buy) limit orders with price p waiting at time t. Further, denote

$$a_t := 0 \vee \min\{p:\ A_t(p) > 0\}, \qquad b_t := n+1 \wedge \max\{p:\ B_t(p) > 0\}$$

the (best) ask, bid respectively.

The list of possible event causing jumps of Ξ together with their intensities is given in the following table:

rate	description
$\theta_t = \theta(a_t, b_t)$	An arrival of a buy market order, causing A^{a_t} to decrease by one (if the sell limit order book is empty then the arrival of the market order has no effect).
$\kappa_{t,p} = \kappa(a_t, b_t, p)$	An arrival of a sell limit order with limit price $p > b_t$ causing an increase of A^p by one.
$\rho_{t,p} = A_t^p \rho(a_t, b_t, p)$	A cancellation of a pending sell limit order with a limit price p causing a decrease of A^p by one.
$\vartheta_t = \vartheta(a_t, b_t)$	An arrival of a sell market order, causing B^{b_t} to decrease by one (if the buy limit order book is empty then the arrival of the market order has no effect).
$\lambda_{t,p} = \lambda(a_t, b_t, p)$	An arrival of a buy limit order with limit price $p < a_t$ causing an increase of B^p by one.
$\sigma_{t,p} = B_t^p \sigma(a_t, b_t, p)$	A cancellation of a pending buy limit order with a limit price p causing a decrease of B^p by one.

Here, $\theta, \kappa, \rho, \vartheta, \lambda, \sigma$ are some functions. It is assumed that all the flows of the market orders, the flows of limit orders and their cancellation are mutually independent in the sense that the conditional distribution of a relative jump time at any fixed time given the history of X up to t is exponential with parameter

$$\Lambda_t = t_t + \vartheta_t + \sum_{p=b_t+1}^{n} \kappa_{p,t} + \sum_{p=0}^{a_t-1} \lambda_{p,t} + \sum_{p=a_t}^{n} A_t^p \rho_{p,t} + \sum_{p=0}^{b_t} B_t^p \sigma_{p,t}$$

and the probability that the next event will be a a particular one equals to ϱ_t/Λ_t where ϱ_t is the event's intensity. It is obvious that Ξ is then a Markov chain in a countable state space.

The following table shows how some of the models, mentioned in the Introduction, comply with our setting.

model	θ	ϑ	κ	ρ	λ	σ
Luckock (2003)	$K(b_t)$	$1 - L(a_t - 1)$	κ^l_p	0	λ^l_p	0
Smith et al. (2003)	θ^s	θ^s	κ^s	ρ^s	κ^s	ρ^s
Cont et al. (2010)	θ^c	θ^c	$\kappa^c(p - b_t)$	$\rho^c(p - b_t)$	$\kappa^c(a_t - p)$	$\rho^c(a_t - p)$

Here, $\kappa^l_p = K(p)-K(p-1)$, $\lambda^l_p = L(p)-L(p-1)$ where K, L are (continuous) cumulative distribution functions and κ^c, ρ^c are some functions and the rest of the the symbols are constants. When speaking about Luckock (2003), we have its discretized version (see Šmíd (2012), Sec 3.3) on our mind. When speaking about Smith et al. (2003), we are considering its bounded version (i.e., contrary to Smith et al. (2003) we assume zero arrival intensities for prices less than one and greater than n; since the bounds of our model may be set arbitrarily large, we can approximate Smith et al. (2003) by our ZI model with an arbitrarily high accuracy).

Some of the models from the Introduction were not mentioned in the table: We did not include Maslov (2000) due to potential technical difficulties with its discretization and because, anyway, its discretized version would be very similar to Smith et al. (2003) with $t^s = 0$. The model by Mike and Farmer (2008), on the other hand, was not included because of its complicated cancellation model and because, apart from the cancellations, it is very similar to that of Cont et al. (2010). Finally, we did not include Stigler (1964) because it is a special version of Luckock (2003) (with $K(x) = L(x) = x$).

Proposition 1. *If $\rho(\bullet) > 0$, $\sigma(\bullet) > 0$, $\theta(\bullet) > 0$ and $\vartheta(\bullet) > 0$. then Ξ_t is ergodic.*

Proof. Our proof mimics that of Cont et al. (2010), Proposition 2 verifying the ergodicity of their model by finding a Markov chain in $\mathbb{N}$ dominating the total number of orders with a recurrent zero state; recurrence of the state then proves a recurrence of the zero state of the dominated model which in

turn verifies the ergodicity (see Cont et al. (2010)). Here we put

$$L = \sum_{p=1}^{n} [\max_{a,b} \kappa(a, b, p) + \max_{a,b} \lambda(a, b, p)]$$

and

$$M_i = \min_{a,b} t(a, b) + \min_{a,b} \vartheta(a, b) + i \min_{a,b,p} [\rho(a, b, p) \wedge \sigma(a, b, p)]$$

and argue that a Markov chain S_t may be constructed such that

$$\sum_{p=1}^{n} [|A_t^p| + |B_t^p|] \leq S_t$$

and having intensity matrix $P = (p_{i,j})_{i,j \in \mathbb{N}}$ with all non-diagonal components zero except of $p_{i,i+1} = L, i \geq 0$, and $p_{i,i-1} = M_i, i > 0$. □

Thanks to to the symmetry between the sell and buy order books in our model, it suffices to deal only with the sell order books until the end of the paper.

Denote $t_1, t_2, \ldots$ the jumps of (a, b) (i.e., the two dimensional process of the quotes). Further, denote

$$U_i = \begin{cases} 1 & \text{if } \Delta a_{t_i} > 0, \\ -1 & \text{if } \Delta a_{t_i} < 0, \\ 0 & \text{otherwise} \end{cases} \qquad i \in \mathbb{N},$$

and

$$E_i = \mathbf{1}[\Delta a_{t_i} > 1], \qquad i \in \mathbb{N},$$

an indicator of more than unit jumps.

Note that, due to the unit size of market orders, the ask may jump more than one tick only if no limit orders with the price one tick above the ask are present in the order book at the time of the jump, i.e.,

$$E_i = 1 \Longleftrightarrow M_i = 0, U_i = 1 \qquad M_i = A_{t_{i-1}^-}^{p_i}, \qquad p_i = a_{t_{i-1}} + 1;$$

Before proceeding to the distribution of M_i, note that, for each $i \in \mathbb{N}$,

$$M_i = M_i^0 + M_i^+$$

where M_i^0 is the number of the orders, present in the book already at the time t_{i-1} which were not cancelled until t_i and M_i^+ is the number of the orders, newly arrived and not cancelled between t_{i-1} and t_i.

Proposition 2.

$$M_i^0|\Xi_{t_{i-1}}, t_i, U_i, U_{i-1} \sim \mathrm{Bi}\left(m_i, \delta(a_{t_{i-1}}, b_{t_{i-1}}, \Delta t_i)\right),$$

$$\delta(a, b, t; \kappa, \rho) = \exp\{-\rho(a+1, a, b)t\}, \qquad m_i = A_{t_{i-1}}^{p_i},$$

and, if $\rho(\bullet) > 0$ *then*

$$M_i^+|\Xi_{t_{i-1}}, t_i, U_i, U_{i-1} \sim \mathrm{Po}\left(\gamma(a_{t_{i-1}}, b_{t_{i-1}}, \Delta t_i)\right)$$

$$\gamma(a, b, t; \kappa, \rho) = \frac{\kappa(a+1, a, b)}{\rho(a+1, a, b)}\left(1 - \delta(a, b, t; \kappa, \rho)\right).$$

Moreover, M_i^0 *and* M_i^+ *are conditionally independent given* $\Xi_{t_{i-1}}, t_i, U_i, U_{i-1}$.

Proof. Note first that, from the Markov property, both M_i^0 and M_i^+ are conditionally independent of U_{i-1} given $\Xi_{t_{i-1}}$ hence it need not be considered when speaking about the distribution of the M's given $\Xi_{t_{i-1}}$ and that, thanks to the strong Markov property (Theorem 12.14 of Kallenberg (2002)), we can, without loss of generality, assume $\Xi_{t_{i-1}}$ to be deterministic. Further, note that we would not change the distribution of Ξ up to t_i if we assume that $\Xi_{t_{i-1}+s} = \Xi'_s$, $0 \leq s < \Delta t_i$, where $\Xi'_0 = \Xi_{t_{i-1}}$ and A'^i_s, B'^i_s, $s > 0$, $1 \leq i \leq n$, are independent immigration and death precesses with the corresponding intensities determined by $\Xi_{t_{i-1}}$, independent also on the first market order arrival time and type. Now, as there are two possible causes of a jump of a up - a market order arrival and a cancellation of the last order with the ask price - which both are caused by either the order flow at tick $a'_{t_{i-1}}$ or the market order process, we are getting that Δt_i is independent of A'^{p_i}, so the distribution of $A_{\Delta t_i}^{p_i}$ may be computed as if the market order flow and $\Xi'^{a'_0}$, and consequently Δt_i and U_i were deterministic (see Šmíd (2012), Lemma A.1. and the discussion below). Our problem hence reduces to finding the distribution of and immigration and death process with an initial value m, the immigration rate κ and death rate ρ at a fixed time t, which is known to be a convolution of a Binomial distribution with parameters m and $\exp\{-\rho t\}$ and a Poisson distribution with intensity $\frac{\kappa}{\rho}(1-\exp\{-\rho t\})$ (see also Šmíd (2012), p 80, for the derivation of the formula and Proposition 4.2 therein for a more exact proof concerning a more general model). $\square$

As a direct consequence of the Proposition, we are getting:

Corollary 1. *Under the assumptions of Proposition 1,*

$$\mathbb{P}(E_i = 1 | \Xi_{t_{i-1}}, t_i, U_i, U_{i-1}) = \omega(a_{t_{i-1}}, b_{t_{i-1}}, \Delta t_i, m_i; \kappa, \rho)$$

$$\omega(a, b, t, m; \kappa, \rho) = \exp\{-\gamma(a, b, t; \kappa, \rho)\} [1 - \delta(a, b, t; \kappa, \rho)]^m$$

whenever $U_i = 1$.

The last result, in fact, opens a door for estimation of the parameters of the order books by means of L1 data because it enables us to indirectly observe M_i whose distribution depends on the parameters in interest which may be consequently estimated by means of the sample of such E_i, for which m_i is observable, which is exactly in when $U_{i-1} = -1$.

Remark 1. *If* $\kappa(p, a, b) = \tilde{\kappa}(p-b)$ *and* $\rho(p, a, b) = \tilde{\rho}(p-b)$ *for all* a, b, p *where* κ *and* ρ *are completely unknown (as in Cont et al. (2010), for instance) then, theoretically, all* $\kappa(2), \dots, \kappa(n-1)$ *and* $\rho(2), \dots, \rho(n-1)$ *could be estimated*[3] *(note that* Ξ *is ergodic so the number of occurrences of each state goes to infinity). In practice, however, there will not be enough observations for estimation of the* κ*'s and* ρ*'s with higher arguments because the probability of states with a large spread (which would be needed to estimate the intensities for higher values of the arguments) is very low.*

3 Data

We use L1 data, i.e. the history of the process

$$(a_t, A^{a_t}, t, b_t, B_t^{b_t}), \qquad t \geq 0,$$

for three US stocks

- Microsoft (MSFT),
- General Electric (GE)
- Exxon Mobile (XOM)

at three US electronic markets

- NASDAQ,

[3]Note that $\kappa(1)$ and $\rho(1)$ can be estimated directly from the L1 data)

- ARCA (NYSE)
- ISE

for our estimation. The data come from the one year period starting 12/2008 and ending 11/2009. Figure 1 shows basic characteristics of the data.

The reasons for which we do not use the L2 (order book) data are three. The first one is the already mentioned difficult accessibility of the limit order data; the second one is the fact, that the estimation by means of the quotes is indirect and hence bringing less information than using the L2 data, is compensated by the much larger amount of available L1 data. The third advantage of L1 data could be a possible existence of hidden limit orders, which might spoil the information provided by the order data; as it is clear from our treatment, this problem is avoided by using the quotes.

A clear disadvantage of the estimation via L1 data is that only the parts of the order books close to quotes may be estimated this way; however, in very liquid markets, this fact does not harm the estimation of the distribution of price movements and/or price impact (which are often the primary topics of an interest) much, because the influence of the deeper parts of the order books in liquid markets is weak.

4 Empirical Evidence

It follows from the ergodicity of Ξ and Lemma 1 (see Appendix) that the (conditional) probabilities

$$\omega^0(t) := \mathbb{P}(M_1 = 0 | t_1 = t, m_1 = 0, U_i = 1, U_{i-1} = -1)$$

and

$$\omega^+(t) := \mathbb{P}(M_1 = 0 | t_1 = t, m_1 > 0, , U_i = 1, U_{i-1} = -1)$$

viewed as functions of t, may be approximated by step functions defined by corresponding empirical frequencies; more preciously, thanks to the ergodicity,

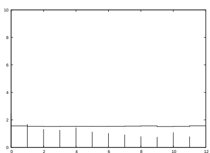
$$\omega^0(t) \doteq \epsilon^0_{\delta,N}(t), \qquad t \geq 0,$$

for $\delta > 0$ small enough and N large enough, where

$$\epsilon^0_{\delta,N}(t) = \sum_{i=1}^{\infty} \mathbf{1}[t \in [(i-1)\delta, i\delta)] u^N_{(i-1)\delta,\delta}, \qquad u^N_{t,\delta} = \frac{\sum_{i=1}^N E_i J^i_{t,\delta}}{\sum_{i=1}^N J^i_{t,\delta}},$$

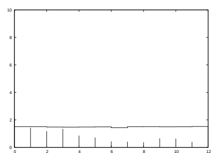
MSFT at Nasdaq

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	1.56	2.39	0.98
1/2009	1.53	3.47	0.67
2/2009	1.52	4.40	0.53
3/2009	1.53	4.64	0.50
4/2009	1.53	4.08	0.57
5/2009	1.53	5.14	0.46
6/2009	1.53	5.65	0.41
7/2009	1.55	6.32	0.37
8/2009	1.56	7.29	0.32
9/2009	1.52	7.72	0.30
10/2009	1.53	5.35	0.44
11/2009	1.57	7.40	0.32



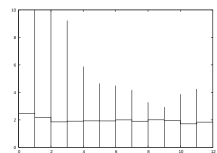
GE at Nasdaq

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	1.51	2.75	0.85
1/2009	1.51	4.12	0.57
2/2009	1.48	4.99	0.47
3/2009	1.48	4.34	0.54
4/2009	1.48	6.84	0.34
5/2009	1.49	8.13	0.29
6/2009	1.43	13.55	0.17
7/2009	1.51	14.24	0.16
8/2009	1.51	15.57	0.15
9/2009	1.50	8.96	0.26
10/2009	1.50	9.21	0.25
11/2009	1.52	14.07	0.16



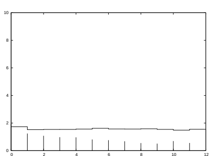
XOM at Nasdaq

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	2.48	0.36	6.55
1/2009	2.18	0.42	5.54
2/2009	1.85	0.56	4.20
3/2009	1.91	0.63	3.69
4/2009	1.93	1.00	2.35
5/2009	1.92	1.26	1.86
6/2009	2.00	1.30	1.80
7/2009	1.90	1.40	1.67
8/2009	2.00	1.78	1.31
9/2009	1.94	1.99	1.17
10/2009	1.71	1.51	1.55
11/2009	1.84	1.35	1.70



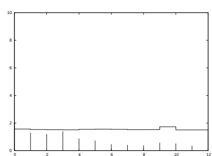
MSFT at ARCA

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	1.73	3.00	0.78
1/2009	1.52	4.72	0.50
2/2009	1.54	5.44	0.43
3/2009	1.54	5.99	0.39
4/2009	1.56	6.06	0.39
5/2009	1.62	7.23	0.32
6/2009	1.57	7.72	0.30
7/2009	1.57	8.60	0.27
8/2009	1.58	10.66	0.22
9/2009	1.54	11.53	0.20
10/2009	1.48	8.44	0.28
11/2009	1.55	10.39	0.22



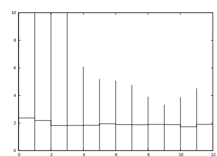
GE at ARCA

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	1.57	2.83	0.82
1/2009	1.53	4.50	0.52
2/2009	1.52	4.89	0.48
3/2009	1.51	4.19	0.56
4/2009	1.54	6.68	0.35
5/2009	1.56	8.01	0.29
6/2009	1.54	12.57	0.19
7/2009	1.52	14.38	0.16
8/2009	1.52	15.26	0.15
9/2009	1.73	10.16	0.23
10/2009	1.50	11.32	0.21
11/2009	1.50	16.38	0.14



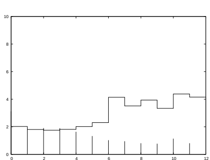
XOM at ARCA

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	2.37	0.34	6.78
1/2009	2.18	0.43	5.47
2/2009	1.82	0.46	5.09
3/2009	1.82	0.53	4.40
4/2009	1.84	0.96	2.43
5/2009	1.95	1.13	2.08
6/2009	1.88	1.15	2.03
7/2009	1.88	1.23	1.90
8/2009	1.90	1.50	1.56
9/2009	1.89	1.76	1.33
10/2009	1.73	1.51	1.55
11/2009	1.90	1.27	1.81



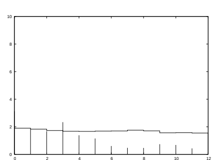
MSFT at ISE

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	2.02	1.83	1.25
1/2009	1.81	3.05	0.77
2/2009	1.75	3.06	0.76
3/2009	1.81	3.08	0.76
4/2009	2.02	3.57	0.65
5/2009	2.30	4.40	0.53
6/2009	4.14	5.68	0.41
7/2009	3.52	6.13	0.38
8/2009	3.94	7.22	0.32
9/2009	3.34	7.53	0.31
10/2009	4.38	5.12	0.46
11/2009	4.15	7.16	0.32



GE at ISE

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	1.90	1.83	1.25
1/2009	1.83	3.19	0.73
2/2009	1.73	3.49	0.67
3/2009	1.68	2.50	0.93
4/2009	1.67	4.22	0.55
5/2009	1.69	5.08	0.46
6/2009	1.70	9.63	0.24
7/2009	1.76	12.30	0.19
8/2009	1.71	12.51	0.19
9/2009	1.56	7.89	0.30
10/2009	1.57	8.56	0.27
11/2009	1.55	13.08	0.17



XOM at ISE

m/y	$\bar{s}$	$\overline{\Delta t}$	#
12/2008	19.50	0.56	4.12
1/2009	10.34	0.55	4.24
2/2009	8.55	0.44	5.35
3/2009	7.30	0.47	5.03
4/2009	4.75	0.61	3.85
5/2009	4.59	0.66	3.57
6/2009	4.09	0.71	3.29
7/2009	5.13	0.76	3.07
8/2009	4.20	0.81	2.90
9/2009	3.22	0.95	2.46
10/2009	3.33	0.77	3.05
11/2009	4.40	0.76	3.01

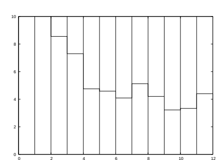
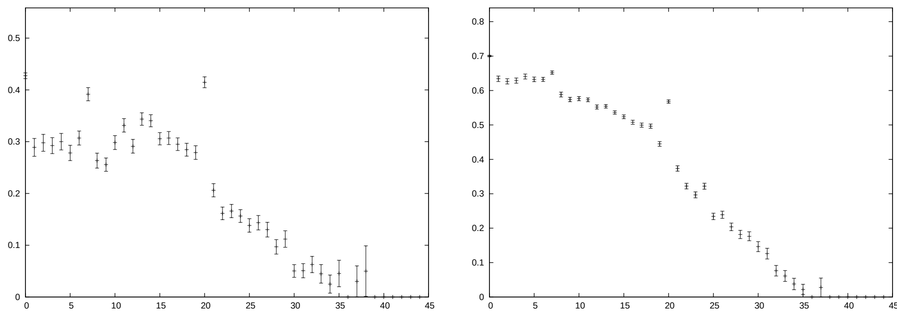


Figure 1: Summary data for the examined stocks and markets. The table: m/y - month/year, $\bar{s}$ - average spread, $\overline{\Delta t_i}$ - average time between jumps of the quotes, # - number of jumps of the quotes a day (in 10000). The graph: vertical lines - number of jumps, the curve - average spread.



(a) $\epsilon^0_{\delta,N}$ against t (b) $\epsilon^+_{\delta,N}$ against t

Figure 2: XOM at ISE - empirical probabilities

$$J^i_{t,\delta} = \mathbf{1}[t_i \in [t, t+\delta), m_i = 0, U_i = 1, U_{i-1} = -1].$$

Function ω^+ may be approximated analogously - denote $\epsilon^+_{\delta,N}$ its approximation. Figure 4 shows such approximations in case of XOM at ISE.

Now, since $\lim_{t\to 0} \omega(a, b, t, 0) = 0$ for each a, b and since $\omega^0(t) = \mathbb{E}_{a_0,b_0}(\omega(a_0, b_0, t, 0))$ it should be $\lim_{t\to 0} \omega^0(t) = 0$ and consequently $\lim_{t\to 0} e^0(t) \doteq 0$. However, the converse is true as indicated by Figure 4 (a) where the empirical version of ω^0 clearly tends to a non-zero value at zero.

Similarly, looking at the approximation of ω^+ (Figure 4 (b)) we are finding that even if it should be approach one at zero (note that $\lim \omega(a, b, t, m) = 1$ for any a, b and $m > 0$), it tends to a clearly non-unit value.

Finally, as, for any a, b, m, $\lim_{t\to\infty} \omega(a, b, t, m) > \iota$ for some ι independent of a, b, m given a non-zero cancellation rate, the limits at ∞ on both (a) and (b) of Figure 4 should be non zero which does not seem to be the case. One may object here that zero limits could be caused by a zero cancellation rate; this would, however, imply $\omega(a, b, t, m) \equiv 0$ for any $m > 0$ (the reason is that the initial orders would never be cancelled so the jump of a could never be non-unit) which is clearly not the case, too.

Summarized, it is clear that the ZI model introduced in Section 2 is too poorly parametrized to explain the empirical data a satisfactory way, because, in reality,

(L0) Limit of $\mathbb{P}(M_i = 0 | m_i = 0, t_i = t)$ as $t \to 0$ is not zero.

(L+) Limit of $\mathbb{P}(M_i = 0 | m_i > 0, t_i = t)$ as $t \to 0$ is not one.

(R) Limit of $\mathbb{P}(M_i = 0|t_i = t)$ as $t \to 0$ is non-zero.

each of which points contradicts the model.

Clearly, our graphs do not falsify the ZI model statistically correct way (nothing was said about the rate of convergence of the estimates of ω). However, these preliminary findings are statistically confirmed in the next Section where a sup-model of the falsified model is shown to be significant in a majority of cases.

Before proceeding further, however, one more discrepancy between the model and reality not discussed yet has to be remembered:

(U) the volumes of the orders are not unit in reality.

5 Generalized Zero Intelligence Model

In the present Section, we introduce a generalization of the ZI model introduced in Section 2 so that it complies with empirical facts (L0), (L+), (R) and (U).

Starting with (L0) note that this in fact says that unexpectedly many limit orders appear during short time periods after jumps of the ask (note that the L1 data cannot contain values of m_i if the ask jumps down more than one tick at t_{i-1}, so we, in accordance with the definition of the model, only assume that $m_i = 0$). We suggest two explanation of those unexpected orders' emergence:

1. If a jumps down more than one tick down at t_{i-1} then not only a new limit order (with limit price equal to the new ask) is put into the spread but some additional one(s) are possibly placed into the ticks between the old and the new value of the ask (perhaps by algorithmic trading) - we assume the numbers of these orders to be Poisson with parameter γ_0.

2. At times of events causing a to jump up (i.e., a market order arrivals and cancellations of the ask) one or more limit orders is (perhaps by algorithmic trading or as a result of a shift of the quote) put one-tick above the original ask at the time of the jump - we assume the number of those orders to be Poisson with parameter η_0.

The fact (L+), on the other hand, indicates that some limit orders expected are missing, which may suggest that a jump of a at t_{i-1} down might be caused not only by a placement of a limit order into the spread but also by a shift of the ask down. Therefore, we assume:

3. At the time of a jump of a down (i.e. a placement of a new order into the spread) each of the orders formerly being the ask might be cancelled with probability α_0

Coming to (R), the most natural explanation this fact seems to the existence of strategic long-term orders, so we suppose:

4. Orders with a negligible cancellation rate ϱ_0 arrive with a rate $\lambda_0 \gg \varrho_0$ into each tick greater than the ask.

Finally, taking (U) into account, we speculate, similarly to e.g., Smith et al. (2003), that the volume of all the orders is $\mu_0 > 0$ instead of one, i.e. that the "actual" market follows a process

$$(\mathbf{A}, \mathbf{B}) = (\mu_0 A, \mu_0 B). \tag{1}$$

As to the arrival and cancellation rates of the generalized model, we assume, similarly to e.g. Mike and Farmer (2008), that $\kappa(a, b, p) = \kappa(p - a)$ and $\rho(a, b, p) = \rho(p - a)$ for each b and $p > a$.

Now, let us define our generalized model so that it takes 1.-4. and (1) into account. Denoting $\tilde{A}^p$ and $\hat{A}^p$ the numbers of short term, long term orders, respectively, and, quite naturally, assuming that any long-term order turns into a short term one if it becomes an ask, we get the list of events in our generalized zero intelligence (GZI) as follows:

event rate	description
$\theta_t = \theta(a_t, b_t)$	An arrival of a buy market order, causing $\tilde{A}^{a_t}$ to decrease by one. If its resulting value is zero then $\tilde{A}^{a_t+1}$ is increased by a random number having $\mathrm{Po}(\eta_0)$ for some $\eta \geq 0$ (the increment is independent of the evolution of the past of the whole process).
$\kappa_{t,p} = \kappa_0$	An arrival of a short-term sell limit order with limit price $p > b_t$ causing an increase of $\tilde{A}^p$ by one. If $p < a_t$ then each order with the price equal to the original ask is cancelled with probability α_0, if the jump of a is more than one tick then each $\tilde{A}^{\varpi}$, $a_{t-} < \varpi < a_t$ is increased by a random number $n_{\varpi} \sim \mathrm{Po}(\gamma_0)$ for some $\gamma_0 \geq 0$ (the random increments and cancellations are mutually independent and independent on the past of the whole process).
$\rho_{t,p} = \tilde{A}^p_t \rho_0$	A cancellation of a pending short term sell limit order with a limit price p causing a decrease of $\tilde{}A^p$ by one. If its resulting value is zero then $\tilde{A}^{a_t+1}$ is increased by a random number having $\mathrm{Po}(\eta_0)$.
$\lambda_{t,p} = \lambda_0$	An arrival of a long term limit order with price $p > a_t$ increasing $\hat{A}^p$ by one.
$\varrho_{t,p} = \hat{A}^p_t \varrho_0$	A cancellation of a long-term order with price p
$\dots$	The definitions concerning sell market orders and buy limit orders are symmetric.

If assume the "independence" analogously to our ZI model, we get that

Proposition 3.

$$\bar{\Xi} = (A, \tilde{A}.B, \tilde{B})$$

is Markov chain with a countable state space.

Moreover,

Proposition 4. *If* $\rho(\bullet) > 0$ *then* $\bar{\Xi}$ *is ergodic.*

Proof. As the increase rate of the total number of the orders is bounded from above (by $2n(\lambda_0 + \eta_0 + \gamma_0 + \max(\kappa(\bullet))))$ while the decrease rate is bounded from below, the total number of limit orders is, similarly to the proof of Proposition 1, bounded from above by an ergodic Markov chain which proves the Proposition. ☐

A brief analysis shows that, now,

$$M_i = M_i^0 + \tilde{M}_i^+ + \hat{M}_i^+ + M_i^{0+} + M_i^{1+}$$

where M_i^0 is the number of uncancelled orders out of those present at t_{i-1}, $\tilde{M}_i^+$ and $\hat{M}_i^+$ are the numbers of uncancelled short-term orders, long-term orders, respectively, out of those having arrived during (t_{i-1}, t_i), M_i^{0+} is the number of (cancellable) orders having appeared at t_{i-1} and M_i^{1+} us the number of the orders having arrived during the jump of a up.

Denote

$$\Omega_i = (\bar{\Xi}_{t_{i-1}}, t_i, U_i).$$

Similarly to the ZI model, we have

$$\tilde{M}_i^+ | \Omega_i \sim \text{Po}\left(\frac{\kappa_0}{\rho_0}(1 - \exp\{-\rho_0 \Delta t_i\})\right),$$

where $\kappa_0 = \kappa(1)$ and $\rho_0 = \rho(1)$, and clearly

$$\hat{M}_i^+ | \Omega_i \sim \text{Po}\left(\frac{\kappa_0}{\varrho_0}(1 - \exp\{-\varrho_0 \Delta t_i\})\right)$$

which we, however, approximate by taking a limit as $\varrho_0 \to 0$ and getting

$$\hat{M}_i^+ | \Omega_i \dot{\sim} \text{Po}(\lambda_0 t),$$

(we did this approximation also because a negligible parameter could hardly be estimated).

Further, from the definition,

$$M_i^{1+} | \Omega_i \sim \text{Po}(\eta_0),$$

and

$$M_i^{0+} | \Omega_i \sim \text{Po}(I(\mathbf{m}_i)\gamma_0 \exp\{-\rho_0 \Delta t_i\}), \qquad I(m) = \mathbf{1}[m = 0]$$

(we have used the fact that if we perform a Bernoulli trial on each "unit" of a Poisson variable, the sum of successes is Poisson with the original intensity multiplied by the Bernoulli probability).

Finally, denoting $\mathbf{m}_i$ the actual volume one tick above the ask, i.e.

$$\mathbf{m}_i = \mu_0 m_i = \mathbf{A}^{p_i}_{t_{i-1}},$$

we have

$$M^0_i|\Omega_i \sim \mathrm{Bi}\left([\mu^{-1}\mathbf{m}_i], (1-\alpha_0)\exp\{-\rho_0 \Delta t_i\}\right).$$

For computational simplicity and a subsequent reduction of the parameter space, we however use the well known approximation $\mathrm{Bi}(n,p) \overset{n\to\infty, p\to 0}{\longrightarrow} \mathrm{Po}(\lim np)$, to get

$$M^0_i|\Omega_i \dot{\sim} \mathrm{Po}\left(\beta_0 \exp\{-\rho_0 \Delta t_i\}\mathbf{m}_i\right), \qquad \beta_0 = \mu_0^{-1}(1-\alpha_0).$$

From our assumptions of independence we now get that

$$\mathbb{P}(E_i = 1|\Omega_i) \doteq \varpi(\mathbf{m}_i, \Delta t_i; \kappa_0, \gamma_0, \lambda_0, \beta_0, \eta_0, \rho_0) \tag{2}$$

$$\begin{aligned}&\varpi(m, t; \kappa, \gamma, \lambda, \beta, \eta, \rho)\\ &\qquad = \exp\left\{-\frac{\kappa}{\rho}\left[1-e^{-\rho t}\right] - \gamma I(m) e^{-\rho t} - \lambda t - m\beta e^{-\rho t} - \eta\right\}\end{aligned}$$

where $\Theta_0 = (\kappa_0, \gamma_0, \lambda_0, \beta_0, \eta_0, \rho_0) \geq 0$ are parameters of our interest.

To estimate Ω_0, we use a non-linear least square estimator

$$\Theta_n = \arg\min_{\Theta \geq 0} S_n(\Theta), \qquad S_n(\Theta) = \sum_{i=1}^{n} [E_i - \varpi(\mathbf{m}_i, \Delta t_i; \Theta)]^2$$

where we put $E_i = \varpi(\mathbf{m}_i, \Delta t_i)) = 0$ by definition whenever $U_i \neq 1$ or $U_{i-1} \neq -1$.[4]

Proposition 5. *If the random element $(E_i, \Omega_i)_{i \in \mathbb{N}}$ is such that the distribution of Ω_i as in the GZI model and the conditional distribution of $(E_i)_{i<n}$ is given by (2) and if the parameter space is bounded from above, then Θ_n is consistent and, if $\rho_0 > 0, \beta_0 > 0$, also asymptotically normal. Moreover, the minimization of NLS is an asymptotically convex problem up to a continuous transformation of parameters.*

[4]Because κ and ρ do not depend on b in the GZI model, we may take t_i as jumps not of the whole (a, b) but of only a.

Proof. See Appendix for all the details. □

Remark 2. *Even if, due to a heteroskedasticity, the weighted version of the NLS estimator suggests itself to be used here, we do not apply it because of a large impact of possible outliers with small predicted residual variance (note that the variance of E_i tends to zero once ϖ is close to zero or to one).*

Remark 3. *Note that we do not deal with the market order arrival rate; the reason for this is that it could be easily estimated directly form the L1 data. Further, as we already said, we leave aside parameters $\kappa(i), \rho(i)$, $i > 1$.*[5]

Remark 4. *Before proceeding to the estimation, let us also note that the asymptotic properties of our estimator would not be lost if we used the exact conditional probability of E_i instead of the approximation.*

6 Empirical Evidence Revisited

In Figure 3, the results of the estimation of the parameters of the GZI model for all the combinations of the three stocks with the three markets, mentioned in Section 3. In case of each stock-market pair, we used a sample of at most 500.000 observations of E_i for such that a jumped up at t_i and a jumped down at t_{i-1}. In addition to the usual statistical analysis including point estimates, standard errors and significance levels, we illustrate the fit of predicted shape of π by its empirical counterparts. In particular, we present four graphs, the k-th one depicting a comparison of a predicted and an empirical version of

$$p_k(\bullet) := \mathbb{P}(E_i = 1 | \Delta t_i = \bullet, \mathbf{m} \in I_k),$$

where $I_1 = \{0\}$ (the top left graph) and I_2, I_3 and I_4 are chosen so that they contain roughly the same number of values of $\mathbf{m}_i$. Similarly, the time axis is distorted so that equal intervals on the x-axis contain comparable numbers of observations of Δt_i. The graphical comparison is an in-sample one.

Naturally, the model fits much better those stock-market pairs which provide larger samples, i.e., those for which the ask jumps more often and by larger magnitudes. If - on the other hand - the number of jumps is too little, the method crashes totally, as it can be seen in case of GE at NASDAQ.

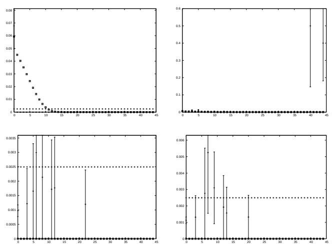
[5] In principle, our methodology could be used to estimate e.g. $\kappa(2)$ and $\rho(2)$ by means of discerning whether the jump of a is by one, two or more ticks, here, however, we omit this for simplicity.

MSFT at Nasdaq

12/2008-11/2009

$\kappa = 8.41(328.82)$, $\gamma = 0.62(0.31)$*,
$\lambda = 17.67(196.69)$, $\beta = 0.24(0.04)$***,
$\eta = 1.97(118.22)$, $\rho = 1.78(10.36)$,

Σ E / n=322/129084,
$\overline{\Delta a} = 1.00$, $\bar{s} = 2.02$

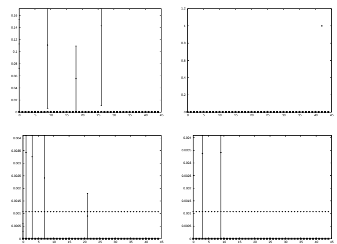


GE at Nasdaq

12/2008-11/2009

$\kappa = 2579.72$(err), $\gamma = 0.00$(err),
$\lambda = 12.78$(err), $\beta = 0.30$(err),
$\eta = 1.68$(err), $\rho = 64.48$(err),

Σ E / n=89/82766,
$\overline{\Delta a} = 1.00$, $\bar{s} = 1.89$

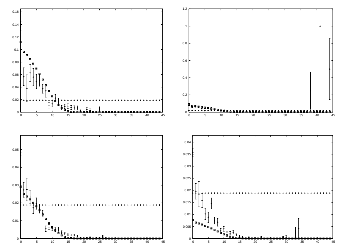


XOM at Nasdaq

12/2008-3/2009

$\kappa = 0.00$(n/a), $\gamma = 0.62$(n/a),
$\lambda = 13.07$(n/a), $\beta = 0.21$(n/a),
$\eta = 1.45$(n/a), $\rho = 0.00$(n/a),

Σ E / n=9419/500000,
$\overline{\Delta a} = 1.03$, $\bar{s} = 2.39$

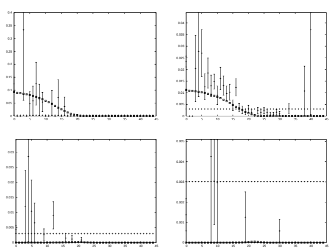


MSFT at ARCA

12/2008-11/2009

$\kappa = 0.83(8.29)$, $\gamma = 0.00(0.24)$,
$\lambda = 3.31(3.50)$, $\beta = 0.10(0.02)$***,
$\eta = 2.33(3.46)$, $\rho = 1.38(1.02)$,

Σ E / n=305/101115,
$\overline{\Delta a} = 1.00$, $\bar{s} = 1.96$

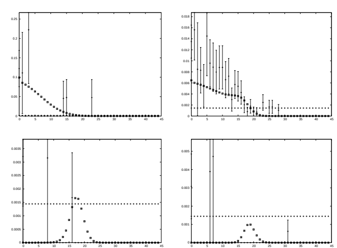


GE at ARCA

12/2008-11/2009

$\kappa = 10.51(5.60)$*, $\gamma = 0.00(0.46)$,
$\lambda = 3.48(1.33)$**, $\beta = 0.15(0.04)$***,
$\eta = 2.19(0.70)$***, $\rho = 6.11(2.06)$**,

Σ E / n=124/85922,
$\overline{\Delta a} = 1.00$, $\bar{s} = 1.96$

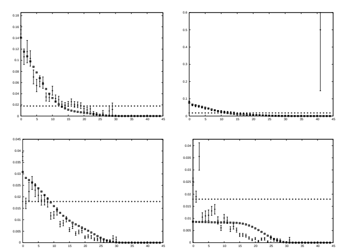


XOM at ARCA

12/2008-3/2009

$\kappa = 19.26(0.47)$***, $\gamma = 0.08(0.05)$,
$\lambda = 0.64(0.07)$***, $\beta = 0.23(0.01)$***,
$\eta = 1.71(0.07)$***, $\rho = 6.79(0.43)$***,

Σ E / n=8977/500000,
$\overline{\Delta a} = 1.02$, $\bar{s} = 2.39$

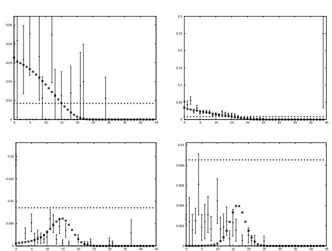


MSFT at ISE

12/2008-11/2009

$\kappa = 20.50(2.86)$***, $\gamma = 2.15(0.17)$***,
$\lambda = 2.98(0.58)$***, $\beta = 0.15(0.01)$***,
$\eta = 1.21(0.29)$***, $\rho = 7.99(1.06)$***,

Σ E / n=1813/212607,
$\overline{\Delta a} = 1.06$, $\bar{s} = 2.17$

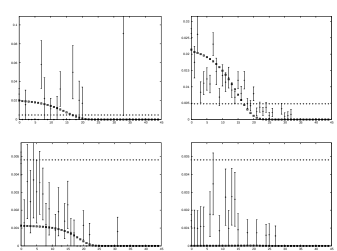


GE at ISE

12/2008-11/2009

$\kappa = 0.09(12.08)$, $\gamma = 1.19(0.21)$***,
$\lambda = 3.80(4.19)$, $\beta = 0.06(0.01)$***,
$\eta = 2.68(6.55)$, $\rho = 0.84(0.70)$,

Σ E / n=834/173372,
$\overline{\Delta a} = 1.02$, $\bar{s} = 2.05$



XOM at ISE

12/2008-5/2009

$\kappa = 0.79(0.03)$***, $\gamma = 0.58(0.01)$***,
$\lambda = 0.17(0.01)$***, $\beta = 0.024(0.00)$***,
$\eta = 0.44(0.01)$***, $\rho = 1.13(0.11)$***,

Σ E / n=232725/500000,
$\overline{\Delta a} = 4.40$, $\bar{s} = 10.23$

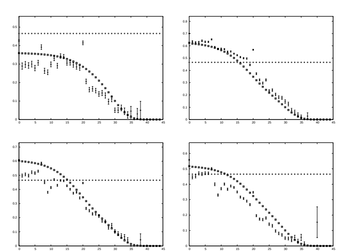


Figure 3: Results of estimation. Table: ΣE - number of jumps of a up greater than 1, n - sample size, $\overline{\Delta a}$ - average jump of a up, $\bar{s}$ - average spread, Graphs: vertical dotted line - average value of E, curved dotted line - predicted shape of p_k, points with errorbars - its empirical version.

Roughly speaking, for the method to be successful, the stock has to be XOM or the market has to be ISE (their combination fitting the best) which leads us to a conclusion that our method is more suitable for moderately liquid stocks/markets than for the super-liquid ones.

7 Conclusion

We have formulated a partially tractable and estimable model of a limit order market which agrees with the data better than any of the previously published zero intelligence, which is illustrated by means of L1 data of three US stocks.

Acknowledgements

This work has been supported by grant No. P402/12/G097 of the Czech Science Foundation.

References

Richard C Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2(107-44):37, 2005.

R. Cont. Statistical modeling of high-frequency financial data. *Signal Processing Magazine, IEEE*, 28(5):16–25, 2011.

R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operations Research*, 56(3):549–563, 2010.

C. Jacob. Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *The Annals of Statistics*, 38(1):566–597, 2010.

O. Kallenberg. *Foundations of Modern Probability.* Springer, New York, second edition, 2002.

H. Luckock. A steady-state model of the continuous double auction. *Quantitative Finance*, 3:385–404, 2003.

S. Maslov. Simple model of a limit order driven market. *Physica A*, 278: 571–578, 2000.

S. Mike and J. D. Farmer. An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics & Control*, 32(1):200–234, 2008.

F. Slanina. Mean-field approximation for a limit order driven market. *Phys. Rev.*, 64:056136, 2001. Preprint at `http://xxx.sissa.it/abs/cond-mat/0104547`.

F. Slanina. Critical comparison of several order-book models for stock-market fluctuations. *The European Physical Journal B-Condensed Matter and Complex Systems*, 61(2):225–240, 2008.

M. Šmíd. Price tails in the smith and farmer's model. *Bulletin of the Czech Econometric Society*, 25, 2008.

M. Šmíd. Probabilistic properties of the continuous double auction. *Kybernetika*, 48(1):50–82, 2012. Available at: http://www.kybernetika.cz/content/2012/1/50/paper.pdf.

E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative finance*, 3(6):481–514, 2003.

G.J. Stigler. Public regulation of the securities markets. *Journal of Business*, pages 117–142, 1964.

A Appendix

First, let us state a nearly obvious theoretical result which we use.

Lemma 1. *If X is a continuous time stationary ergodic Markov chain in countable space then $Y = (\Delta\tau_i, X_{\tau_i})_{i\in\mathbb{N}}$, where τ_i is the i-th jump time, is stationary ergodic stochastic process.*

Proof. Denote $\Lambda(x)$ the intensity of the jump of X given that the state of $X = x$. From the strong Markov property (Theorem 12.14 of Kallenberg (2002)), from Lemma 12.16 of Kallenberg (2002) and from the scalability of exponential distribution we have that $U_1, U_2, \dots, U_i = \Delta\tau_i/\Lambda(X_{\tau_{i-1}})$, $i \in \mathbb{N}$,

is a sequence of i.i.d. (unit exponentially) distributed variables, independent of $X_0, X_{\tau_1}, \dots$. As U_n - being an i.i.d. sequence - is a strong mixing and X_{τ_n} is a strong mixing by Bradley (2005), process $Z_n := (X_{\tau_n}, U_n)$ is a strong mixing (for any "rectangles" $A = A^1 \times A^2$ and $B = B^1 \times B^2$, $\mathbb{P}(Z \in A \cap t_n B) = \mathbb{P}(Z^1 \in (A^1 \cap t_n B^1))\mathbb{P}(Z^2 \in (A^2 \cap t_n B^2)) \to \mathbb{P}(Z^1 \in A^1)\mathbb{P}(Z^1 \in t_n B^1)\mathbb{P}(Z^2 \in A^2)\mathbb{P}(Z^2 \in t_n B^2) = \mathbb{P}(Z \in A)\mathbb{P}(Z \in B)$, the case of general A, B follows from their approximation by rectangles) we get the ergodicity of Z by the well known fact that strong mixing implies ergodicity. Finally, as Y_n is a function of Z_n, the Lemma is proved. □

A.1 The NLS estimator

Before proving the desired properties of our NLS estimator, we perform a reparametrization of the problem simplifying the further theoretical analysis. In particular, we introduce two additional parameters:

$$\phi = \frac{\kappa}{\rho}, \qquad \zeta = \frac{\kappa}{\rho} + \eta.$$

The vector of the parameters is now

$$\Upsilon_0 = (\phi_0, \gamma_0, \lambda_0, \beta_0, \zeta_0, \rho_0)$$

and we have

$$\mathbb{P}(E_i = 1|\Omega_i) \doteq \varpi(\mathbf{m}_i, \Delta t_i; \Upsilon)$$
$$\varpi(m, t; \Theta) = \exp\left\{-\zeta - \gamma I(m)e^{-\rho t} - \lambda t - m\beta e^{-\rho t} + \phi e^{-\rho t}\right\}.$$

A.1.1 Convexity of minimization

By a textbook differentiation

$$\nabla_\Upsilon \varpi(m, t; \Upsilon) = -\delta(m, t; \Upsilon)\varpi(m, t; \Upsilon),$$
$$\delta(m, t; \Upsilon) = \begin{bmatrix} -e^{-\rho t} \\ I(m)e^{-\rho t} \\ t \\ me^{-\rho t} \\ 1 \\ te^{-\rho t}\left[\phi - I(m)\gamma - m\beta\right] \end{bmatrix} = e^{-\rho t} \begin{bmatrix} -1 \\ I(m) \\ e^{\rho t} t \\ m \\ e^{\rho t} \\ t\left[\phi - I(m)\gamma - m\beta\right] \end{bmatrix} \tag{3}$$

and

$$\nabla_{\Upsilon\Upsilon}\varpi(m,t;\Upsilon) = -\varpi(m,t;\Upsilon)[H(m;\Upsilon) + \delta(m,t;\Upsilon)\delta(m,t;\Upsilon)']$$

$$H(m,t;\Upsilon) = te^{-\rho t}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -I(m) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -m \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -I(m) & 0 & -m & 0 & t\left[\phi - I(m)\gamma - m\beta\right] \end{vmatrix} \tag{4}$$

Hence for

$$s(E,t,m;\Upsilon) = [\varpi(m,t;\Upsilon) - E]^2$$

we have

$$\nabla_{\Upsilon}s = 2\nabla_{\Upsilon}\varpi[\varpi - E]$$

and

$$\begin{aligned} \nabla_{\Upsilon\Upsilon}s &= 2(\nabla_{\Upsilon}\varpi)(\nabla_{\Upsilon}\varpi)' + 2(\varpi - E)\nabla_{\Upsilon\Upsilon}\varpi \\ &= 2\varpi^2\delta\delta' - 2\varpi(\varpi - E)(te^{-\rho t}H + \delta\delta') \\ &= 2\varpi(E - \varpi)te^{-\rho t}H + 2\varpi E\delta\delta' \end{aligned}$$

So

$$\begin{aligned} \nabla_{\Upsilon\Upsilon}S = \frac{2}{n}\sum_{i=1}^{n}&[\varpi(\mathbf{m}_i,\Delta t_i;\Upsilon)(E_i - \varpi(\mathbf{m}_i,\Delta t_i;\Upsilon)\Delta t_i e^{-\rho\Delta t_i})H(\mathbf{m}_i;\Upsilon)] \\ &+ \frac{2}{n}\sum_{i=1}^{n}[\varpi(\mathbf{m}_i,\Delta t_i;\Upsilon)E_i\delta(\mathbf{m}_i,\Delta t_i;\Upsilon)\delta(\mathbf{m}_i,\Delta t_i;\Upsilon)'] \end{aligned}$$

Since, by the Lemma 1 and Birkhof theorem, the left hand sum goes to zero as $n \to \infty$ and the convergence is uniform if we bound parameters ϕ, γ, β from above (which is assumed), the minimization of S is asymptotically convex problem (note that $\delta\delta' \geq 0$).

A.1.2 Consistency and Asymptotic Normality

We use the theory Jacob (2010) to prove the asymptotic properties of our estimator. First, note that $f_k, \eta_k, \mathbf{f}'_k, \mathbf{f}''_k, d_k$ and M_n from Jacob (2010) evaluate, in our case, as

$$f_k(\Upsilon) = \varpi(\mathbf{m}_k, \Delta t_k; \Upsilon),$$

$$
\begin{aligned}
\eta_k &= E_k - f_k(\Upsilon) \\
\mathbf{f}'_k(\Upsilon) &= -\varpi(\mathbf{m}_k, \Delta t_k; \Upsilon)\delta(\mathbf{m}_k, \Delta t_k; \Upsilon) \\
\mathbf{f}''_k(\Upsilon) &= \varpi(\mathbf{m}_k, \Delta t_k; \Upsilon)\left[\delta(\mathbf{m}_k, \Delta t_k; \Upsilon)\delta(\mathbf{m}_k, \Delta t_k; \Upsilon)' + H(\mathbf{m}_k, \Delta t_k; \Upsilon)\right] \\
d_k(\Upsilon) &= \varpi(\mathbf{m}_k, \Delta t_k; \Upsilon_0) - \varpi(\mathbf{m}_k, \Delta t_k; \Upsilon), \\
M_n &= \left[\sum_{k=1}^{n} \varpi^2(\mathbf{m}_k, \Delta t_k; \Upsilon)\delta(\mathbf{m}_k, \Delta t_k; \Upsilon)\delta'(\mathbf{m}_k, \Delta t_k; \Upsilon)\right]^{-1}
\end{aligned}
$$

First let us verify the conditions, required for the consistency:

- $\mathrm{LIP}_\Upsilon(\{\varpi(\mathbf{m}_\mathrm{k}, \Delta \mathrm{t}_\mathrm{k}; \Upsilon)\})$: From the assumed boundedness of Υ and from (3) it follows that the derivative of ϖ is uniformly bounded by a non-random constant B so the condition is fulfilled with $g_k \equiv 1$ and $h(x) = B\|x\|$.

- VAR_Υ: As $\mathrm{var}(E_j|\Omega_i) \leq 1$, the condition follows from Jacob (2010), Proposition 5.2.

- $\mathrm{SI}_\Upsilon(\{D_n(\Upsilon)\})$: It follows from (3) and the non-negativeness of ϕ_0 that, for all $1 \leq j \leq 6$, $\frac{\partial}{\partial \Upsilon_j} d(0, t; \Upsilon) \neq 0$ for any $t > 0$. From the continuity of the derivatives it further follows that, for any t_0, $r_1 > 0, \ldots, r_6 > 0$ exist such that $|\partial \Upsilon_j d(0, t; \Upsilon)| \geq r_j$ for any Υ from the parametric space (recall it is bounded) from which it follows that $|d_k(\Upsilon)| \geq \min(r_1, \ldots, r_6)\|\Upsilon - \Upsilon_0\|$ whenever $\Delta t_k \leq t_0$. Hence, for any δ,

$$
\inf_{\|\Upsilon - \Upsilon_0\| \geq \delta} \sum_{i=1}^{N} d_k(\Upsilon)^2 \geq K_N \min(r_1, \ldots, r_6)\delta
$$

  where $K_N = |\{k \leq N : \mathbf{m}_k = 0, \Delta t_k \leq t_0\}|$. Thanks to the ergodicity of $\bar{\Xi}$ and Lemma 1, $K_N \to \infty$, so the condition is proved.

Thanks to the conditions and Proposition 3.1. of Jacob (2010) the strong consistency of Υ_n is proved, implying the same for Θ_n.

As to the asymptotic normality, note first that, from the ergodicity,

$$
\frac{1}{n} M_n^{-1} \to M^- \tag{5}
$$

$$
M^- = \mathbb{E}_{Q(m,t)}[\varpi^2(m, t; \Upsilon)\delta(m, t; \Upsilon)\delta'(m, t; \Upsilon)]
$$

where Q is a stationary distribution of the process $(\mathbf{m}_i, \Delta t_i)$.

Lemma 2. *If $\rho \neq 0$ and $\beta \neq 0$ then M^- is regular.*

Proof. We show that $M^- > 0$, i.e.

$$v'M^-v > 0, \qquad v \neq 0,$$

which suffices for the regularity. As

$$M^- = \mathbb{E}dd', \qquad d = \varpi(m,t;\Upsilon)\delta(m,t;\Upsilon), \qquad (m,t) \sim Q$$

which clearly implies that $M^- \geq 0$, it suffices to show that $d'v \neq 0$ on a set with a positive probability for any non-random vector v: in particular, as $\varpi > 0$ everywhere, it suffices to show that

$$r(m,t) > 0, \quad \text{with a non-zero probability}, \qquad r(m,t) = v'\delta(m,t;\Upsilon)$$

which we do by contradiction: Assuming a converse we get

$$\sum_{j=0}^{\infty} p_j \int \mathbf{1}[r(j,t) \neq 0] g(t|j) dt = 0$$

where g is a density of t given m and p_j is a probability that $m = j$. As the distribution of t is a mixture of exponential distributions, $g(\bullet|m)$ is continuous non-zero, and, as m is a mixture of Poisson distributions, $p_j > 0$, $j \in \mathbb{N}$; therefore, for each $j \in \mathbb{N}$ there should exist a non-trivial interval J_j such that

$$r(j,t) = 0, \qquad t \in J_j, j \in \mathbb{N}$$

which would, for each j, imply the existence of t_j and $\Delta_j >$ such that

$$\frac{\partial}{\partial_s} r(j, t_j + s) = 0 \quad s \in [-\Delta_j, \Delta_j]$$

yielding

$$e^{-\rho t_j} e^{-\rho s}(v_1 - v_2 I(j) - v_4 j - v_6 c_j(t_t + s - 1)) + v_3 \equiv 0, \qquad s \in [-\Delta_j, \Delta_j]$$

where $c_j = \phi - I(j)\gamma + j\beta$, which could happen only if

$$v_3 = 0, \qquad v_1 - v_2 I(j) - v_4 c_j = 0, \qquad v_6 c_j = 0, \qquad j \in \mathbb{N}.$$

From the regularity of matrix $(1, I(j), c_j)_{j=0}^{3}$, we are getting $v_1 = v_2 = v_3 = 0$. Moreover, since at least one $c_j, j \in \mathbb{N}$, is non-zero, we have also $v_6 = 0$ which implies $v = (0,0,0,0,v_5,0)$, $v_5 \neq 0$. However, since $v'\delta(m,t;\Upsilon) = v_5^2$, we are getting $v_5 = 0$ which is a contradiction to non-triviality of v. $\square$

Consequetnly, it follows from the continuity of the inversion operator that also

$$nM_n \to M, \qquad M := (M^-)^{-1} \tag{6}$$

Now let us verify the conditions needed by Jacob (2010) to get the asymptotic normality:

- $\mathrm{LIP}_t(\mathbf{f}_{k;i,l})$: Follows from the twice-differentiability
- VAR_Υ was discussed when proving consistence
- $\mathrm{UNC}(\{\Upsilon_n\})$ follows from (6)
- $\mathrm{SI}(\{(M_n[i,j])^{-1}\})$: Since $M^- > 0$, it has to be $M > 0$ implying positiveness of its components which proves, together with (6) the convergence rate of M_n to be exactly n^{-1} from which SI follows easily.
- $\mathrm{LIM}(\{\Upsilon_n\})$: follows from the convergence of M_n to zero and boundedness of the derivatives.

Now, put

$$\xi_{k,n} = n^{-\frac{1}{2}} \mathbf{f}'_k(\Upsilon_0)\eta_k$$

Clearly, $\xi_{k,n}$ is a triangular array of martingale differences with

$$\lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}(\xi_{i,n}\xi_{i,n}^T | \xi_{1,n}, \ldots, \xi_{i-1,n}) = \lim_{n\to\infty} \frac{1}{n} A_n = A$$

$$A_n = \sum_{i=1}^{n} a_i, \qquad a_i = \varpi_i(\Upsilon_0)(1-\varpi_i(\Upsilon_0))\mathbf{f}'_k(\Upsilon_0)\mathbf{f}'^T_k(\Upsilon_0)$$

$$A = \mathbb{E}(\varpi(m,t;\Upsilon_0)(1-\varpi(m,t;\Upsilon_0))\mathbf{f}'(m,t;\Upsilon_0)\mathbf{f}'^T(m,t;\Upsilon_0))$$

by the ergodicity. Therefore and because, for each $\epsilon > 0$,

$$\sum_{i=1}^{n} \mathbb{E}(\|\xi_{i,n}\|^2 \mathbf{1}_{[\|\xi_{i,n}\|\geq\epsilon]} | \xi_{1,n}, \ldots, \xi_{i-1,n}) = \frac{1}{n} \sum_{i=1}^{n} \|a_i\|^2 \mathbf{1}_{[\frac{1}{n}\|a_i\|\geq\epsilon]} \to 0$$

(recall that $|\eta_i| \leq 1$), we can use the CLT for multidimensional martingale difference arrays (as cited by Jacob (2010) in Section 8) to get

$$\sum_i \xi_{i,n} \to \mathcal{N}(0, A)$$

Consequently, putting,
$$\Psi_n = n^{-\frac{1}{2}} M_n^{-1}$$
(note that M_n has to be regular starting from certain n) and using (5) and (6) we finally get that

$$\Psi_n M_n \sum_i \xi_{i,n} \to \mathcal{N}(0, A)$$

which gives, by Proposition 6.1 of Jacob (2010), that

$$n^{-\frac{1}{2}} M_n^{-1} (\Upsilon_n - \Upsilon_0) \to \mathcal{N}(0, A)$$

in distribution so, from the (5), also

$$n^{\frac{1}{2}} M^{-} (\Upsilon_n - \Upsilon_0) \to \mathcal{N}(0, A)$$

yielding
$$n^{\frac{1}{2}} (\Upsilon_n - \Upsilon_0) \to \mathcal{N}(0, MAM).$$
Now, by using the Continuous mapping theorem, we get

$$n^{\frac{1}{2}} (\Theta_n - \Theta_0) \to \mathcal{N}(0, TMAMT^T).$$

where
$$T = \begin{vmatrix} \rho_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$
allowing us to approximate

$$\Theta_n - \Theta_0 \dot{\sim} \mathcal{N}(0, TM_nA_nM_nT^T).$$